

## PCOS Detection using Hybrid CNN-XG Boost Model - A Multimodal Data Approach

Shital Pawar, Pratik Dhane, Dhanraj Shelke\*,  
Prashant Dheple and Nisarg Doshi

Department of Electronics and Telecommunication Engineering,  
Vishwakarma Institute of Technology Pune, India.

\*Corresponding Author E-mail: dhanrajshelke03@gmail.com

<https://dx.doi.org/10.13005/bpj/3252>

(Received: 16 March 2025; accepted: 08 July 2025)

Polycystic ovarian syndrome (PCOS) affects 8-13% of women of reproductive age, with up to 70% of cases being untreated globally. In the proposed methodology, medical images of ovarian ultrasounds are processed through VGGNet-19, extracting high-level features from the images. Simultaneously, clinical text reports, containing valuable diagnostic information, are subjected to a zero-shot learning text classification model. This hybrid architecture enables the utilization of both visual and textual data sources for enhanced PCOS detection. The fusion model leverages VGGNet-19's prowess in image feature extraction, capturing intricate patterns and details within ultrasound images. A dataset of size 2004 images is used for this experiment. XGBoost, known for its robust classification capabilities, processes the extracted features to classify PCOS cases effectively. The proposed system uses both textual and image data for early PCOS detection. The textual data has been made using various factors, including dietary habits, daily routines, and more. The proposed system has the capability for early detection of PCOS while simultaneously identifying concurrent health conditions associated with PCOS like diabetes hypertension etc. The dataset encompasses various parameters, including patient-specific information such as age, weight, height, and BMI, as well as medical measurements like Hb, cycle characteristics, hormone levels, and more. The deep fusion model achieved an accuracy of 99.6% .

**Keywords:** CNN; Disease Detection; Polycystic Ovary Syndrome (PCOS); XGBoost; Zero-Shot Learning.

Polycystic ovarian syndrome, or PCOS, is a common hormonal disorder that affects up to 20% of reproductive-aged women. Among its many symptoms are irregular periods, acne, excessive hair growth, and difficulties with reproduction. Diagnosing PCOS can be difficult as it requires a blend of clinical and laboratory assessments, despite its widespread occurrence and substantial impact on women's well-being. Current PCOS diagnostic methods like the Rotterdam consensus

criteria can be complex and costly. There's a demand for a non-invasive approach to make early diagnosis and treatment more accessible. Recently, machine learning algorithms have shown promise as early PCOS diagnosis and screening methods.

The research introduces an innovative approach to predict PCOS in women by integrating both image and textual data into the diagnostic process. Images are processed through a Convolutional Neural Network (CNN) known as

VGGNet-19 network to extract features describe the physical characteristics of the ovaries and any potential cysts. These image features are then classified using the XGBoost algorithm, which is adept at handling structured data. Concurrently, textual information, such as medical records and patient history, is analyzed using a Zero-Shot learning methodology, enabling the model to make predictions based on text that may not explicitly mention PCOS-related terms. The results from both image and text analyses are combined to provide a comprehensive diagnosis, potentially facilitating earlier detection and more effective treatment of PCOS in women of childbearing age. This combined technique marks a substantial leap in PCOS diagnosis.

This approach represents a promising advancement in the field of PCOS diagnosis, as it leverages both visual and textual data, potentially leading to earlier detection and more effective treatment of this common endocrine disorder in women of childbearing age.

#### **Literature Review**

This study looks at the application of machine learning algorithms to diagnose female polycystic ovarian syndrome (PCOS). The suggested model is an ensemble model made up of the Random Forest and XGBoost algorithms. 177 PCOS patients and 541 women make up the dataset that was analyzed. PCOS is predicted using a number of ML algorithms, including Ada boost, Random Forest, XG boost, Decision tree, and a hybrid model. The suggested approach, which combines XGBoost and Random Forest, detects PCOS with 97.2% accuracy. The treatment provided to women with PCOS may be enhanced by early identification of the illness.<sup>1</sup>

The common symptoms of PCOS, which can make getting pregnant difficult, are listed in the article. These symptoms include excessive weight gain, acne, hair loss, skin discoloration, facial hair growth, and irregular periods.<sup>2</sup> It emphasizes how important it is to detect PCOS early and treat it in order to reduce the risk of long-term issues including heart issues and type 2 diabetes. The method proposed in this study uses ML algorithms that use a limited number of factors to forecast how successful PCOS medication will be discusses the use of two distinct datasets for classification: one uses CNN to categorize photos as normal or PCOS

other involves identifying the key traits underlying PCOS and forecasting the disease using various machine learning methods.

The application of ML is the study's main focus methods to classify PCOS from ovarian ultrasound images. The paper's main objective is to classify women according to their PCOS status without the need for medical intervention.<sup>3</sup> This is accomplished by the analysis of ovulation ultrasound pictures to detect the existence of Polycystic Ovarian Morphology (PCOM). In order to improve the PCOM outcomes, the article trains the pictures using a huge number of datasets. techniques for classification like SVM, K-NN, and (CNN) are employed, with the CNN algorithm achieving an accuracy of 90%.

Regular exercise is beneficial for treating PCOS as it helps lower cholesterol levels, decrease insulin resistance, and fight obesity. PCOS increases the risk of endometrial cancer in females. This paper proposes the use of the YOLO algorithm (You Only Look Once) for the automatic detection of cysts in sonography images of PCOS.<sup>4</sup>

Python programming is used to propose and build CNN-based methods for the categorization of cysts in ultrasound pictures. The study uses CNN-based image process with key-point extract to classify the cysts in the dataset. The study makes use of an independent trained dataset of disorders associated with PCOS.<sup>5</sup> The feature extraction process is performed on a test dataset, resulting in 85% accuracy using performance factors. The suggested approach classifies diseases using an effective CNN-based image processing model.

The study suggests a unique method for zero-shot object detection that uses class semantics to synthesize visual features for unseen classes.<sup>6</sup> The goal of this study is to learn both visible and unseen things in the visual domain, whereas the current zero-shot detection methods concentrate on mapping visual properties to semantic domains for observed objects. The proposed generative model not only generates features for unseen objects but also discriminatively separates them, ensuring high diversity and fluctuating localisation accuracy in the bounding boxes that were found.

The study proposes a method for the early diagnosis and prediction of PCOS using machine learning techniques such as logistic regression KNN, classification and regression trees (CART),

random forest classifier (RFC), and support vector machines (SVM).<sup>7</sup> The performance of several current algorithms must be compared in order to design a suitable machine-learning model for PCOS diagnosis. Following the selection of the most crucial elements from the clinical and metabolic test results using SPSS V 22.0, the feature set is modified using Principal Component Analysis (PCA). When the model's performance is evaluated in terms of accuracy, with an accuracy of 89.02%, the RFC is demonstrated to be the most appropriate and accurate method for PCOS prediction. Though no specifics are given, the system's development includes using efficient tools and accessible platforms.

The study employed Extreme Gradient Boosting (XGBoost), a probabilistic ensemble-based gradient boosting technique, for PCOS early diagnosis. The authors combined SMOTE (Synthetic Minority Oversampling Technique) with other methods to resample their data.<sup>8</sup>

The proposed optimized-SVLR hybrid model, Op-RMSprop, showed efficient performance in predicting PCOS with an accuracy of 89.03%. The hybrid methodology, which combines the SVM linear kernel with Logistic Regression functionalities, along with the RMSprop optimizer, proved to be effective in analyzing PCOS datasets and improving the model's output. In terms of accuracy, the Op-RMSprop model fared better than other machine learning algorithms currently in use, including SVM, Logistic Regression, Decision Tree, KNN, RFC, and Adaboost.<sup>9</sup>

The paper reviews research on the diagnosis of diseases using different ML architectures and specifically focusses on Polycystic Ovary Syndrome (PCOS) diagnosis. Faster R-CNN, a deep learning model, has not been used yet for the diagnosis of PCOS,<sup>10</sup> but it has the potential to detect smaller objects and lead to more accurate and timely diagnosis. Other object detection algorithms such as YOLO, CNN, SPP-Net, R-CNN, and fast R-CNN can also be used for the detection of cysts in ovaries.

## MATERIALS AND METHODS

### Dataset

The model prediction was based on utilizing two distinct datasets: one comprised of

ultrasonic images. At the same time, the other contained textual data encompassing diverse factors associated with the daily lives of various women.

### Dataset Description

#### Ultrasonic image dataset

The ultrasonic image dataset comprises a collection of 574 diverse images specifically focusing on the ovaries. These images are instrumental in the context of polycystic ovarian syndrome (PCOS) as they offer crucial insights into ovarian morphology, including the presence of cysts or other structural abnormalities. Such comprehensive imaging information is essential for PCOS diagnosis, tracking, and comprehension.

#### Textual dataset

The dataset with 43 rows and 542 columns used for PCOS prediction is a comprehensive dataset aimed at studying and predicting polycystic ovarian syndrome (PCOS). Each row likely represents a different individual, while each of the 542 columns likely contains various features, attributes, or variables related to those individuals' health, medical history, lifestyle, and other relevant factors.

Figure 1 depicts the process flow diagram of the proposed system used for identification of PCOSA.

### Algorithm

**Input:** Image and a textual data regarding the current medical condition

**Output :** PCOS detection along with other medical conditions

**Step 1:** Read the input image and use **VGGNet-16** to extract high-level features from ovarian ultrasound images, capturing intricate patterns and details.

**Step 2:** **XGBoost** processes the extracted features to effectively classify PCOS cases.

**Step 3: Text Embedding** - The clinical text reports are subjected to a zero-shot learning text classification model, which enables the utilization of valuable diagnostic information from the textual data.

**Step 4: The SoftMax function** is applied to calculate probabilities for PCOS classification based on the combined information from image features and clinical text data

**Step 5: The BERT model is fine-tuned** for PCOS classification by minimizing the objective function through gradient descent and backpropagation.

## METHODOLOGY

### Data Preprocessing

To facilitate extracting meaningful information from a diverse dataset of images, a crucial aspect lies in preparing the data for subsequent analysis. In this context, a series of preprocessing steps were applied to the dataset. Labels were extracted by parsing the directory structure, capturing the label information associated with each image. To ensure consistency and ease of analysis, images were resized to a uniform dimension of (256, 256) using OpenCV's resize function. Furthermore, to ensure that the images were uniformly represented, grayscale images were transformed into three-channel images using the cv2.merge function. Importantly, a stringent check for image validity was implemented to avoid processing empty or corrupted image files. Finally, the preprocessed images and their associated labels were stored in separate arrays, ready to be utilized for subsequent experiments. This preprocessing pipeline ensures that the dataset is amenable to a wide range of computer vision and machine learning techniques, providing a solid foundation for training the further model.

In the proposed methodology amalgamation of two distinct datasets, each encompassing essential patient information is done. The first dataset primarily comprised hormonal profiles, while the second encompassed a range of clinical features. The initial step involved data integration, ensuring seamless blending of these diverse datasets. Various features within the dataset underwent thorough analysis, with a specific focus on discerning markers for diseases such as diabetes. The proposed methodology involved a comprehensive assessment of patient profiles, including physiological attributes like age, weight, and blood pressure, alongside hormonal markers and lifestyle factors. This textual dataset has two classes "Yes" indicating PCOS and "No" indicating there is no PCOS in the given individual.

### Visualization of data

Figure 2 showcases the image types used from both classes, providing a visual representation

of the dataset. Additionally, Figure 3 offers a bar graph illustrating the count of patients within both classes utilized during Zero-Shot training. These visual representations contribute to a comprehensive understanding of the data collection and training process in this research paper.

### Features Extraction

VGGNet-19, a CNN architecture, is utilized for image feature extraction. Given an ultrasound image represented as  $I$ , VGGNet-19 processes it through a series of convolutional layers followed by max-pooling layers. Figure 4 depicts the convolutional layers used and their output shape and the output of the last convolutional layer is denoted as  $F_p$ , a high-dimensional feature vector:

$$F_{VGG} = \text{VGGNet-19}(I) \quad \dots(1)$$

Where:

- 1). *VGGNet-19* represents the VGGNet-19 model.
- 2).  $F_{VGG}$  is a feature vector of size  $N$  (number of features).

### Classification of Ultrasound Images with XGBoost

Once the high-level features are extracted  $F_{VGG}$ , we use the XGBoost algorithm for effective classification of PCOS cases. XGBoost is renowned for its resilience and capacity to manage intricate categorisation assignments. The classification process is represented by Equation No. 2 as:-

$$y \text{XGBoost} = \text{XGBoost}(F_{VGG}) \quad \dots(2)$$

### Integration with Zero-Shot Learning

To improve the model's generalization across various clinical text reports, zero-shot learning is employed to classify PCOS-related patterns even when they haven't been explicitly encountered during training. The zero-shot learning process involves.

### Text Embedding with BERT

Clinical text reports are embedded using BERT, a cutting-edge pre-trained transformer-based model for natural language understanding. Mathematically, the text embedding  $E_{\text{report}}$  for a clinical text report is obtained as:

$$E_{\text{report}} = \text{BERT}(\text{Clinical Text}) \quad \dots(3)$$

where BERT processes the clinical text

to produce embeddings. BERT helps understand the connections in text data and can make various predictions, like identifying early signs of issues such as diabetes and high blood pressure, using zero-shot learning. This combination of advanced technology and early health insights represents a significant advancement in health problem

prediction. The BERT framework understands the context and relationships in the provided text data through its pre-trained knowledge of language patterns and associations. When making early-stage predictions, it relies on the textual factors in the dataset rather than specific training examples. The power of the model is in its capacity to forecast

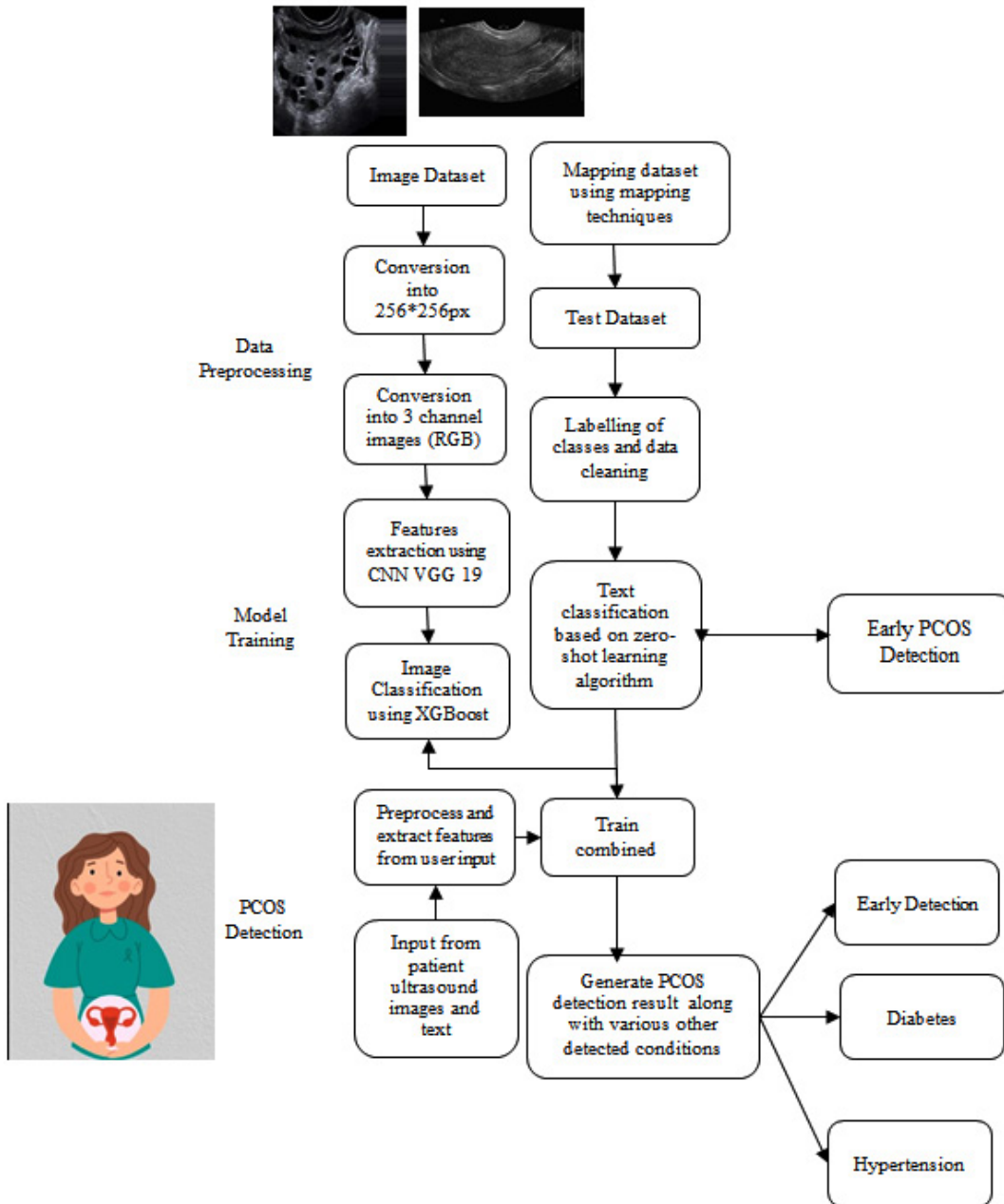


Fig. 1. Project Flow Diagram for the proposed multimodal system

novel, unobserved situations by extrapolating from its training data. It learns to grasp the language and context of the data, enabling it to draw conclusions even in cases where it hasn't seen identical examples during training. This adaptability makes it particularly valuable for tasks like early-stage

prediction, where comprehensive training data may be limited.

**Class-Attribute Compatibility**

The similarity between the report embedding and class attributes  $A_i$  is assessed using cosine similarity:

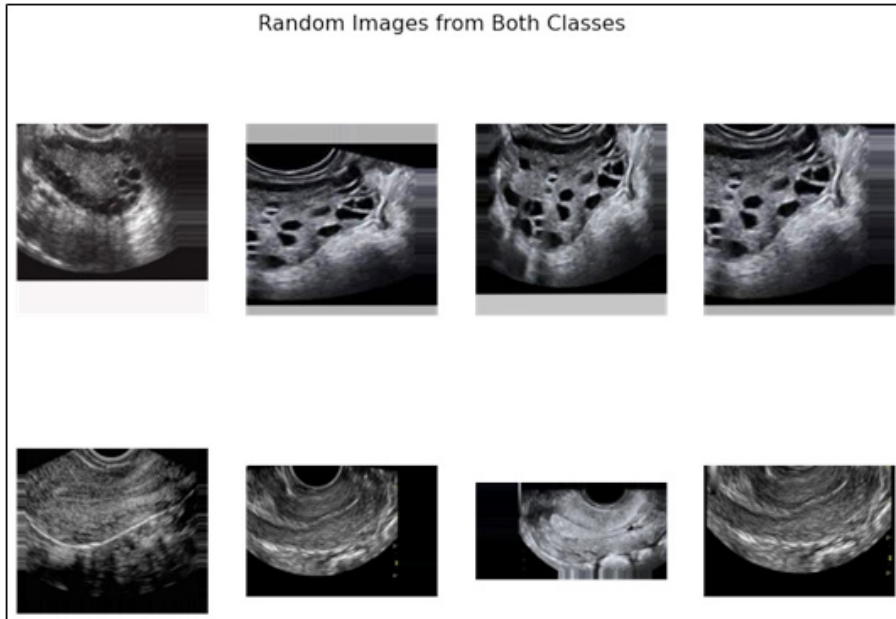


Fig. 2. Image depicting the type of images used from both classes for training the model

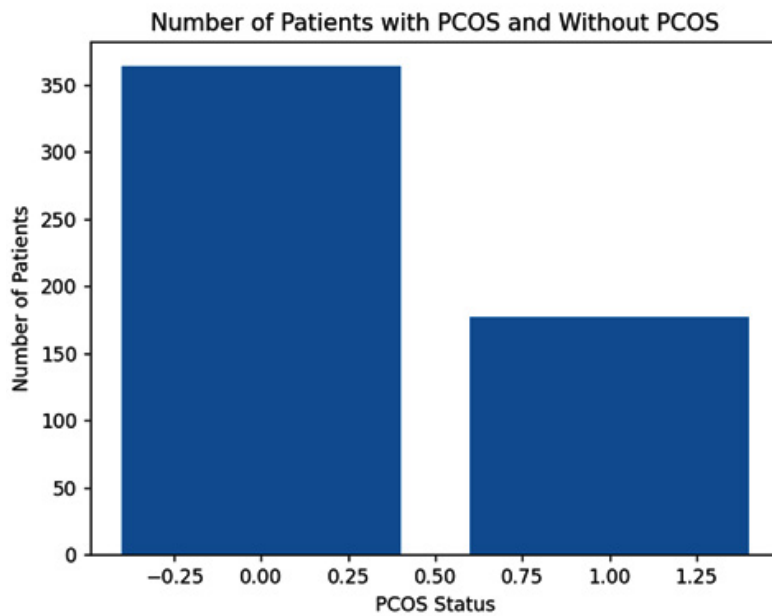


Fig. 3. Bar Graph Depicting the count of patients for both classes used in Zero-Shot training

$$S(E_{\text{report}}) = \text{Cosine Similarity}(E_{\text{report}}, A_i) \dots(4)$$

This measures how compatible the report is with each class attribute.

**Probability Calculation**

The SoftMax function is used to calculate the probability of the report belonging to each class, represented by Equation No. 5 as

```

Model: "vgg19"
-----
Layer (type)                Output Shape                Param #
-----
input_1 (InputLayer)        [(None, 256, 256, 3)]      0
block1_conv1 (Conv2D)       (None, 256, 256, 64)       1792
block1_conv2 (Conv2D)       (None, 256, 256, 64)       36928
block1_pool (MaxPooling2D)  (None, 128, 128, 64)      0
block2_conv1 (Conv2D)       (None, 128, 128, 128)     73856
block2_conv2 (Conv2D)       (None, 128, 128, 128)     147584
block2_pool (MaxPooling2D)  (None, 64, 64, 128)       0
block3_conv1 (Conv2D)       (None, 64, 64, 256)       295168
block3_conv2 (Conv2D)       (None, 64, 64, 256)       590080
block3_conv3 (Conv2D)       (None, 64, 64, 256)       590080
block3_conv4 (Conv2D)       (None, 64, 64, 256)       590080
block3_pool (MaxPooling2D)  (None, 32, 32, 256)       0
block4_conv1 (Conv2D)       (None, 32, 32, 512)       1180160
block4_conv2 (Conv2D)       (None, 32, 32, 512)       2359808
block4_conv3 (Conv2D)       (None, 32, 32, 512)       2359808
block4_conv4 (Conv2D)       (None, 32, 32, 512)       2359808
block4_pool (MaxPooling2D)  (None, 16, 16, 512)       0
block5_conv1 (Conv2D)       (None, 16, 16, 512)       2359808
block5_conv2 (Conv2D)       (None, 16, 16, 512)       2359808
block5_conv3 (Conv2D)       (None, 16, 16, 512)       2359808
block5_conv4 (Conv2D)       (None, 16, 16, 512)       2359808
block5_pool (MaxPooling2D)  (None, 8, 8, 512)         0
-----
Total params: 20,024,384
Trainable params: 0
Non-trainable params: 20,024,384
    
```

**Fig. 4.** Figure depicting convolutional layers used in vgg19

$$P(y_{zero-shot} = i) = \frac{e^{s(E,A_i)}}{\sum_{j=1}^N e^{s(Ereport,A_j)}} \dots(5)$$

Where N: total No. of classes.

**Fine-Tuning of BERT Model**

The process of fine-tuning a pre-trained BERT model for PCOS classification entails modifying the model’s weights and settings to guarantee peak performance on the particular job that these approaches recommend. This procedure is essential to the model’s ability to comprehend and categorize PCOS-related clinical text reports. When fine-tuning, the model parameters  $\theta$  are updated in connection to the PCOS classification task in order to minimize an objective function  $J(\theta)$ . The objective function typically involves a combination of the pre-training loss  $J(\theta)$  and the classification loss  $J(\theta)$  represented by Equation No. 6:

$$J(\theta) = J_{report}(\theta) + J_{classification}(\theta) \dots(6)$$

The objective function is minimized by using the gradient descent technique. The model’s parameters  $\theta$  are iteratively updated in the opposite direction as the gradient with respect to  $\theta$ , as shown by Equation No. 7:

$$\Theta \leftarrow \theta - \eta \nabla_{\theta} J(\theta) \dots(7)$$

where  $\eta$  represents the learning rate, controlling the step size in the parameter updates.

Backpropagation is used to compute the gradient  $\nabla_{\theta} J(\theta)$ , which is the gradient of the loss function  $J(\theta)$  with respect to the parameters of each layer in the BERT model. The weights and biases of the model are modified using this gradient.

The arrangement in the PCOS classification problem, the dissimilarity between the genuine labels and the predicted class probabilities is measured by the  $J(\theta)$  loss. It frequently gets computed using an appropriate loss function,

**Table 1.** Table depicting the classification report for the deep fusion model used for the classification of images

| precision    | recall | f1-score | support |
|--------------|--------|----------|---------|
| infected     | 1.00   | 1.00     | 1.00    |
| not infected | 1.00   | 1.00     | 1.00    |
| accuracy     |        | 1.00     | 1823    |
| macro avg    | 1.00   | 1.00     | 1.00    |
| weighted avg | 1.00   | 1.00     | 1.00    |

**Table 2.** Evaluation of zero-shot learning model

|               |                             |
|---------------|-----------------------------|
| Epoch 1/5     | Validation Accuracy: 0.7481 |
| Epoch 2/5     | Validation Accuracy: 0.7832 |
| Epoch 3/5     | Validation Accuracy: 0.7867 |
| Epoch 4/5     | Validation Accuracy: 0.8456 |
| Epoch 5/5     | Validation Accuracy: 0.8621 |
| Test Accuracy | 0.8732                      |

like cross-entropy, which is shown by Equation No. 8.:

$$J_{classification}(\square) = \frac{-1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log \log (p_{ij}) \dots(8)$$

where  $y_{ij}$  is an indicator variable that indicates whether example  $N$  is the number of training instances,  $C$  is the number of classes (in this case, PCOS-related and non-PCOS), and  $I$

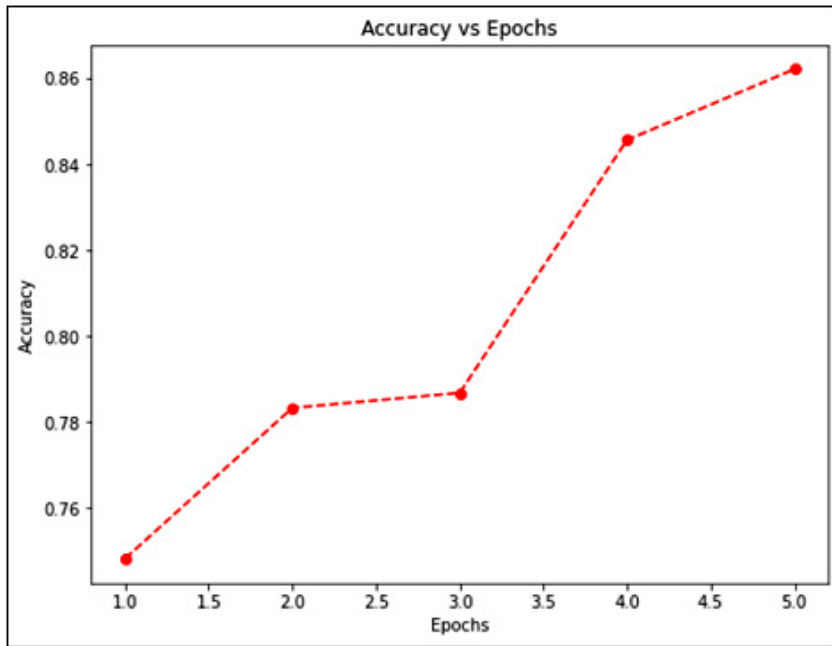


Fig. 5. Graph depicting epochs vs accuracy for the zero-shot model

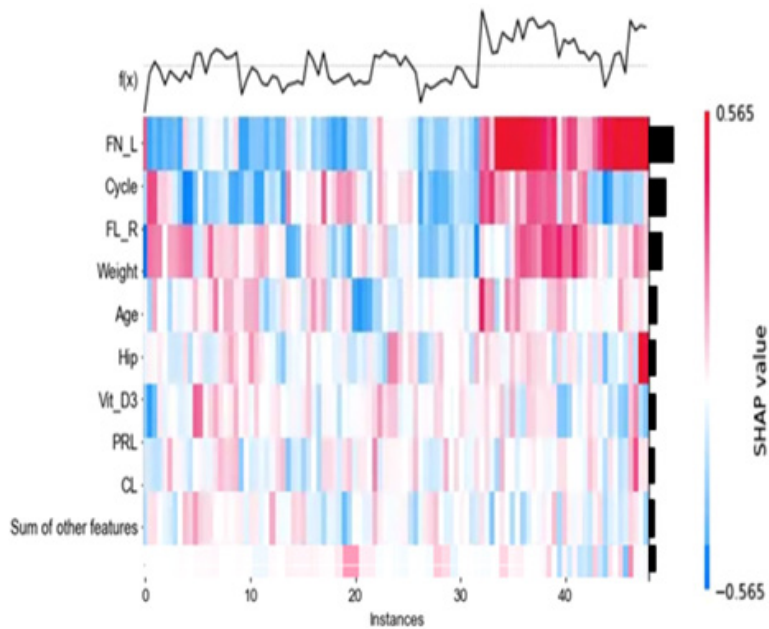


Fig. 6. Heat Map of developed Model

**Table 3.** Comparison of the existing model with the proposed system

| Model   | Accuracy |
|---|----------|
| Ada boost, Random Forest, XG boost, Decision tree, and a hybrid model                                   | 97.20%   |
| SVM , K-NN and CNN  | 99.00%   |
| CNN   | 85.00%   |
| SVM, Random Forest Classifier, KNN, CART, logistic regression, and the Naïve Bayes classifier technique | 89.02%   |
| Extreme Gradient Boosting, XGBoost,VGG16  | 99.20%   |
| Multimodal PCOS Detection   | 99.60%   |

belongs to class  $j$ .  $P_{ij}$  is the projected probability that an example of  $I$  belongs to class  $j$ .

Pre-training and classification losses are combined to create an objective function that must be optimized to fine-tune the BERT model for PCOS classification. To minimize this objective function, the model's parameters are modified via gradient descent and backpropagation. By modifying the model's representations, this procedure enables it to effectively categorize PCOS-related data in clinical text reports.

#### Final Prediction

The probability distribution serves as the foundation for Zero-Shot's final estimation. By using zero-shot learning, the suggested model is more competent in identifying and categorizing PCOS-related patterns that were not visible during training, improving diagnostic precision and resilience.

#### Ensemble of BERT Models

In the text classification model, employed an ensemble of BERT models, each pre-trained on distinct corpora and domains. Denoting these pre-trained models as  $M_1, M_2$  where  $k$  indicates No. of BERT models in ensemble.

The outcome of each BERT model is a text embedding:

$$E_{\text{BERT}} = M_i(\text{Clinical Text}), \quad (9)$$

For  $i=1,2$

Where,

M1: - Hugging Face BERT Transformer

M2: - RoBERTa

#### Combining Model Outputs

To create a robust and adaptable text classification model, combining the outputs of these BERT models using an attention mechanism. Let  $\hat{\alpha}_i$  represent the attention weight assigned to

the  $i$ -th BERT model's output. The combined embedding is obtained, which is depicted in Figure 6:

$$E_{\text{combined}} = \sum_{i=1}^k \hat{\alpha}_i E_{\text{BERT}} \quad \dots(10)$$

#### Classification with Zero-Shot Learning

Then, zero-shot learning, as described above, is applied to classify clinical text reports based on the compatibility between the combined embedding and class attributes. This ensemble approach allows the model to capture a broader spectrum of language patterns and domain-specific knowledge, making it highly adaptable to unseen classes during inference. These advanced techniques, when integrated into the fusion model, provide a powerful framework for PCOS detection that combines the strengths of image-based and text-based data, enhancing diagnostic accuracy and robustness.

## RESULTS

The efficacy of the multimodal PCOS detection model was assessed using common measures, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). Cross-validation experiments were also conducted to determine ability of the model to generalise to fresh data.

Table 1 shows various values like precision, recall, f1-score, and support for the image classification model.

99.6% accuracy was demonstrated by the deep fusion model when the model was tested on a test dataset which is depicted in Figure 7 of the ROC curve. The zero-shot learning capabilities empowered the model to discern rare

or atypical PCOS presentations. Through the exploitation of semantic relationships between textual descriptions and well-established PCOS symptoms, the model effectively identified and flagged unconventional cases. This functionality carries broader implications beyond PCOS and extends to other medical conditions characterized by diverse and infrequent manifestations. Such adaptability in recognizing uncommon medical scenarios underscores the versatility and potential applicability of zero-shot learning in the realm of medical diagnosis and beyond. Table 3 shows the accuracy achieved by the zero-shot model in 5 epochs.

## DISCUSSION

The proposed multimodal PCOS detection model demonstrated a high sensitivity for identifying PCOS in its early stages. This was evident from a significant increase in true positive predictions for cases where PCOS was in its incipient phase. Early-stage sensitivity is a key factor in preventing the progression of PCOS-related complications. The model's proficiency in early-stage prediction facilitates timely interventions and personalized treatment plans. Healthcare professionals can leverage these predictions to initiate appropriate therapies and lifestyle modifications, potentially halting the progression of the condition before it leads to more severe health issues.

One of the study's main conclusions is that PCOS diagnosis ability is much enhanced when picture and textual data are combined. The XGBoost model effectively captures image patterns related to PCOS, while zero-shot learning enables the model to understand textual clinical reports, leading to a more comprehensive understanding of the patient's condition.

### Existing Model Comparison

In the experiments, a comparison was made between the performance of the multimodal approach and single-modal models, focusing solely on image data and textual data. The outcomes revealed that the multimodal model consistently outperformed both single-modal models. Notably, the accuracy of the multimodal model exhibited a substantial increase compared to that of the individual models, emphasizing the pivotal role of

integrating image and textual data for enhancing PCOS detection capabilities.

The multimodal model presented in the paper emphasizes early-stage detection, which is crucial for managing various medical conditions. By targeting the disorder in its initial phases, the model aims to make it easier to manage or treat. A significant strength of this model is its ability to predict not only the primary disorder but also related diseases. For example, if the primary disorder is a neurodegenerative condition, the model can potentially predict secondary health problems commonly linked to it, such as diabetes, hypertension, or cardiovascular issues. This comprehensive approach allows healthcare professionals to address multiple health concerns simultaneously, improving patient care and overall wellness.

## CONCLUSION

The proposed PCOS detection system effectively combines image analysis and text processing techniques. It exhibits high accuracy, robustness, and generalization capabilities, making it a valuable tool for diagnosing PCOS based on ultrasound images and clinical text reports. With an impressive 99.6% accuracy rate on the test dataset, it demonstrates robustness and adaptability. By leveraging high-level features from ultrasound images and clinical text reports, this system offers a powerful tool for accurate PCOS diagnosis, potentially improving early detection and patient care in healthcare settings. By offering a dependable and effective technique for early PCOS screening, this research advances healthcare technology, ultimately benefiting patients and healthcare professionals.

## ACKNOWLEDGEMENT

The author would like to thank Vishwakarma Institute of Technology, Pune for granting the B.Tech. research work. The Department of Electronics and Telecommunication, Vishwakarma Institute of Technology, Pune is highly appreciated for allowing the Computer laboratory work.

**Funding Sources**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**Conflict of interest**

The authors do not have any conflict of interest.

**Data Availability Statement**

This statement does not apply to this article.

**Ethics Statement**

This research did not involve human participants, animal subjects, or any material that requires ethical approval.

**Informed Consent Statement**

This study did not involve human participants, and therefore, informed consent was not required.

**Clinical Trial Registration**

This research does not involve any clinical trials.

**Permission to reproduce material from other sources**

Not Applicable.

**Authors Contribution**

Pratik Dhane: Medical image processing, feature extraction (VGGNet-19); Dhanraj Shelke: XGBoost model development, hybrid model optimization; Prashant Dheple: Literature review, ML algorithm comparative analysis; Nisarg Doshi: Zero-Shot learning for clinical text classification.

**REFERENCES**

1. Tanvi, Jain. Detection of PCOS using Ensemble Models. *International Journal For Science Technology And Engineering*: 2023;45-53.
2. V, Srinithi., R., Rekha. Machine learning for diagnosis of polycystic ovarian syndrome (PCOS/PCOD). *International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS)*: 2023;19-24.
3. Rashmita, Khilar. Classification of PCOS Using Machine Learning Algorithms Based on Ultrasound Images of Ovaries. *Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*: 2023;1-7.
4. K. Mahajan and P. Mane. Follicle Detection of Polycystic Ovarian Syndrome (Pcos) Using Yolo. *9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India*: 2023;246-257.
5. M Sumathi, P Chitra, R Sakthi Prabha and K Srilatha. Study and detection of PCOS related diseases using CNN. *IOP Conf. Ser.: Mater. Sci. Eng.* 1070 012062 :2021;764-775
6. Nasir, Hayat., Munawar, Hayat., Shafin, Rahman, et al. Synthesizing the Unseen for Zero-Shot Object Detection. *Asian Conference on Computer Vision 2020*: 2020;155-170.
7. A. Denny, A. Raj, A. Ashok, C. M. Ram and R. George i-HOPE: Detection And Prediction System For Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques. *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India*: 2019;673-678.
8. M. S. Khan Inan, R. E. Ulfath, F. I. Alam, F. K. Bappee and R. Hasan. Improved Sampling and Feature Selection to Support Extreme Gradient Boosting For PCOS Diagnosis. *IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), NV, USA*: 2021;453-467.
9. H. Batra, K. Saluja, S. Gupta, R. Kaushal, N. Sharma and P. Singh. Machine Learning Techniques for Data-Driven Computer-Aided Diagnostic Method of Polycystic Ovary Syndrome (PCOS) resulting from Functional Ovarian Hyperandrogenism (FOH). *International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India*: 2023; 195-201.
10. R. P, Naveen. Op-RMSprop (Optimized-Root Mean Square Propagation) Classification for Prediction of Polycystic Ovary Syndrome (PCOS) using Hybrid Machine Learning Technique. *Jan -International Journal of Advanced Computer Science and Applications*: 2022; 13-26.